

JC05 Rec'd PCT/PTO 16 SEP 2005

10/549661

## 명세서

단일염기 다형을 포함하는 대장암과 관련된  
 폴리뉴클레오티드, 그를 포함하는 마이크로어레이 및 진단  
 키트 및 그를 이용한 대장암의 진단방법{A  
**POLYNUCLEOTIDE ASSOCIATED WITH A COLON  
 CANCER COMPRISING SINGLE NUCLEOTIDE  
 POLYMORPHISM, MICROARRAY AND DIAGNOSTIC  
 KIT COMPRISING THE SAME AND METHOD FOR  
 DIAGNOSING A COLON CANCER USING THE  
 POLYNUCLEOTIDE}**

## 기술분야

- [1] 본 발명은 대장암 관련된 용도의 폴리뉴클레오티드, 이를 포함하는 마이크로어레이, 진단 키트 및 대장암의 진단방법에 관한 것이다.

## 배경기술

- [2] 모든 생물의 계통은 진화의 과정에서 자손의 서열에 변이체를 생기게 하는 자발적 돌연변이를 겪는다 (Gusella, Ann. Rev. Biochem. 55, 831-854(1986)). 변이체는 선조 형태에 비하여 진화적으로 이익 또는 불이익을 주거나 중성적일 수 있다. 어떤 경우는 변이체가 치사적 불이익을 주어 자손에게 전달되지 않는 경우도 있다. 다른 경우에는, 종에게 진화학적인 이익을 주고, 결국에는 종의 대부분에 DNA가 삽입되어 효과적으로 선조 형태가 된다. 많은 경우 이 선조 형태 및 변이체는 살아남아 종집단 중에 공존하게 된다. 서열의 복수 형태의 공존에 의하여, 다형 (polymorphism)이 발생한다.
- [3] 이러한 다형에는 RFLP (restriction fragment length polymorphism: 제한 단편 길이 다형), STR (short tandem repeats), VNTR (variable number tandem repeat), SNP(single nucleotide polymorphism : 단일염기 다형) 등이 알려져 있다. 이중 SNP는 단일염기 다형으로서, 동일한 종의 개체 사이의 단일뉴클레오티드 변이의 형태를 취한다. 단일염기 다형은 코딩 영역 서열에 발생하는 경우, 다형 형태 중 어느 하나에 의하여 결함 있거나 변이된 단백질이 발현될 수도 있다. 다른 경우에는 비코딩 영역 서열에 단일염기 다형이 발생할 수 있다. 이들 중 일부는 결함 있거나 변이된 단백질의 발현을 초래할 수 있다 (예를 들면, 결함 있는 스플라이싱의 결과로서). 어떤 단일염기 다형은 표현형에 아무런 영향을 미치지 않을 수 있다.

- [4] 단일염기 다형은 인간의 경우 약 1,000bp 마다 1회의 빈도로 발생하는 것으로 알려져 있다. 이들 단일염기 다형이 질병과 같은 표현형에 영향을 미치는 경우, 상기 단일 염기 다형을 포함하는 폴리뉴클레오티드는 이러한 질병을 진단하는 데에 프라이머 또는 프로브로서 사용될 수 있다. 상기 단일염기 다형에 특이적으로 결합하는 모노클로날 항체 또한, 질병의 진단에 사용될 수 있다. 현재 여러 연구 기관에서 단일염기 분석 및 그 기능 분석에 관한 연구를 수행하고 있다. 이렇게 찾아낸 단일염기 다형의 염기서열 및 기타 실험결과를 데이터베이스화하여 공개, 누구나 접근하여 연구에 이용할 수 있도록 하고 있다.
- [5] 그러나 이러한 단일염기 다형은 단순히 인간의 게놈 또는 cDNA 상에 단일염기 다형이 존재한다는 것만을 발견하였을 뿐, 이들이 표현형에 미치는 영향을 밝힌 것은 아니었다. 이들 중 일부에 대하여는 그 기능이 알려진 것도 있으나, 대부분이 알려지지 않았다.
- [6] 대장암은 우리 나라를 포함한 전 세계에서 매우 흔하게 발생하는 암이다. 우리 나라에서는 남녀 모두 4번째로 흔하게 발생하는 암으로, 암에 의한 사망률도 4 번째로 인구 십만 명당 약 7명이 대장암으로 사망하고 있다. 최근 10년 사이 사망률도 약 80% 정도 증가하는 추세에 있다.
- [7] 대장암의 발생에는 환경인자의 비중이 크다고 알려져 있으며, 주 원인으로는 식생활의 급격한 서구화, 동물성 지방 또는 단백질의 과다 섭취를 들 수 있다. 그러나, 5% 전후의 대장암은 유전적 소인에 의해 발생한다고 알려져 있다.
- [8] 환자의 90% 이상이 40세 이상에서 발생하고 있는데, 나이와 더불어 대장암 관련 가족력, 염증성 장 질환, 대장 용종 (확인 요망), 난소암, 자궁암, 유방암을 앓았던 경우 더 자주 발생한다 (high-risk group)고 알려져 있다. 30-40대의 젊은 사람에게서 발생하는 경우에는 유전적 소인에 의한 경우가 많다.
- [9] 대장암은 조기에 발견할 경우 100%에 가까운 완치률을 보이지만, 일반적으로 특징적인 자각증상이 없기 때문에 조기 발견에 어려움이 있다. 무증상 시기에 대장 암을 발견하기 위해서 선별검사인 잠혈검사 (대변 내 미량의 출혈도 확인할 수 있는 검사)를 통해 정밀검사를 받을 필요가 있는 사람을 골라내는 방법을 사용하고 있다. 그러나, 이 방법은 위양성 (false-positive), 및 위음성 (false-negative) 비율이 높기 때문에 결과를 조기진단에 사용하기에는 적합하지 않다. 현재 대장암의 진단확정을 위해서는 대장조영검사, 대장내시경, 방사선 진단 등의 방법을 사용하고 있다. CEA라는 종양표지자가 일반적으로 이용되나, 대장암의 진행도와 치료 효과의 판정에 지표로 사용될 뿐, 혈액검사를 통해 대장암을 조기에 발견하거나 예측할 수 있는, 검증된 종양표지자 (marker)는 아직 없는 것으로 알려져 있다. 대장암에 걸리기 쉬운 고위험군 (high-risk group)의 환자를 골라내거나 조기 진단하기 위한 마커들은 몇몇 알려져 있으나,

이들을 이용할 경우 대다수의 대장암 환자들을 골라내는 데는 적합하지 않은 단점이 있다.

- [10] 종래 대장암을 비롯한 각종 암 및 복잡한 질병을 진단 혹은 조기 예측하는데 가장 큰 문제점은 이들 질환은 어느 정도 진행되어야 물리적인 방법을 사용하여 진단해 낼 수 있다는 것이다. 그러나, 최근의 여러 가지 분자생물학적 기술의 발전과 인간유전체에 대한 연구가 1차적으로 완료됨에 따라 질병과 직접 또는 간접적으로 관련있는 유전자 혹은 유전적 변이를 찾아내어 질병의 조기진단에 사용할 수 있게 됨으로써, 기존의 표현형 혹은 외형적인 질병의 특징에 의존하던 진단법으로부터 유전적 요인을 가지고 질병의 발병 여부를 예측할 수 있는 조기진단이 가능하게 되었다. 기존에 사용되고 있는 대장암의 진단 방법에는 생화학적 및 분자생물학적 방법 등이 있다. 그 중 분자생물학적 방법에 의한 대장암 진단의 경우, 질병과 관련 있는 유전자 또는 유전적 변이에 대한 정보 및 대장암 발병률과의 관계 등에 대한 정보의 부족으로 환자나 전문의가 원하는 수준의 조기진단은 이루어지고 있지 못하고 있다. 또한, 하나의 생물학적 마커를 진단에 이용하는 대부분의 경우 마커의 민감도 (sensitivity)와 특이도 (specificity) 모두가 높은 값을 가지는 경우는 드물고, 민감도가 높은 경우 특이도가 낮거나 그 반대의 경우가 대부분이다. 따라서, 진단 시 오류가 발생할 가능성이 높기 때문에 전문의가 원하는 수준의 정확성을 확보하지 못하는 경우가 많아, 대부분 정밀진단을 위한 사전 스크리닝 개념의 진단 마커로 이용되고 있다.

#### 발명의 상세한 설명

##### 기술적 과제

- [11] 본 발명의 목적은 대장암과 관련 있는 단일염기 다형을 포함하는 폴리뉴클레오티드를 제공하는 것이다.
- [12] 본 발명의 또 다른 목적은 상기 대장암과 관련있는 단일염기 다형을 포함하는 폴리뉴클레오티드를 포함하는 마이크로어레이 또는 대장암 진단용 키트를 제공하는 것이다.
- [13] 본 발명의 또 다른 목적은 본 발명의 대장암과 관련 있는 폴리뉴클레오티드를 이용하여 대장암을 진단하는 방법을 제공한다.

##### 기술적 해결방법

- [14] 본 발명은 서열번호 1 내지 12의 폴리뉴클레오티드로 구성된 군으로부터 유래한 10개 이상의 연속 뉴클레오티드를 포함하는 폴리뉴클레오티드로서, 101번 위치의 다형성 부위 (n) 염기를 포함하는 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드를 제공한다.
- [15] 상기 폴리뉴클레오티드의 길이는 서열번호 1 내지 12의 다형성 부위 (101번 위치 : 각각 "n"으로 표시하였다. 이하 동일)를 포함하는 10개 이상의

뉴클레오티드를 포함하는 것이다. 상기 길이는 10 내지 200 뉴클레오티드, 바람직하게는 10 내지 100 뉴클레오티드, 더욱 바람직하게는 10 내지 50 뉴클레오티드인 것이다.

[16] 상기 서열번호 1 내지 12의 뉴클레오티드 서열을 갖는 폴리뉴클레오티드는 다형성 서열이다. 상기 다형성 서열 (polymorphic sequence)은 뉴클레오티드 서열 중에 단일염기 다형을 나타내는 다형성 부위 (polymorphic site)를 포함하는 서열을 말한다. 다형성 부위 (polymorphic site)란 다형성 서열 중 단일염기 다형이 일어나는 부위를 말한다. 상기 폴리뉴클레오티드는 DNA 또는 RNA가 될 수 있다.

[17] 본 발명에 있어서 상기 서열번호 1 내지 12의 101번 위치의 다형성 서열 중 다형성 부위 (n)는 대장암과 연관성이 있는 것이다. 이는 대장암 환자와 정상인으로부터 유래한 혈액 시료로부터 얻어진 DNA를 서열분석한 결과를 통하여 확인할 수 있었다. 그 결과를 다음 표 1 및 2에 나타내었다.

[18] 표 1. 서열번호 1 내지 12의 다형성 서열과 대장암과의 연관성

[19]

| ASSAY_ID | SNP   | 서열<br>번호 | 대립인자빈도 |        |       | 유전자형 출현수     |              |              |              |              |              |
|----------|-------|----------|--------|--------|-------|--------------|--------------|--------------|--------------|--------------|--------------|
|          |       |          | cas_A2 | con_A2 | Delta | cas_<br>A1A1 | cas_<br>A1A2 | cas_<br>A2A2 | con_<br>A1A1 | con_<br>A1A2 | con_<br>A2A2 |
| CCK048   | [A/C] | 1        | 0.945  | 0.973  | 0.028 | 0            | 25           | 204          | 0            | 16           | 277          |
| CCK061   | [A/G] | 2        | 0.646  | 0.714  | 0.068 | 31           | 101          | 98           | 22           | 120          | 145          |
| CCK117   | [A/C] | 3        | 0.636  | 0.555  | 0.081 | 24           | 107          | 82           | 51           | 157          | 83           |
| CCK162   | [G/C] | 4        | 0.647  | 0.714  | 0.067 | 31           | 99           | 98           | 21           | 120          | 142          |
| CCY_041  | [T/G] | 5        | 0.61   | 0.507  | 0.103 | 32           | 109          | 81           | 67           | 140          | 71           |
| CCY_056  | [A/T] | 6        | 0.39   | 0.299  | 0.091 | 106          | 60           | 57           | 144          | 120          | 27           |
| CCY_065  | [T/G] | 7        | 0.409  | 0.328  | 0.081 | 84           | 105          | 42           | 132          | 126          | 32           |
| CCY_067  | [A/C] | 8        | 0.377  | 0.286  | 0.091 | 97           | 85           | 42           | 147          | 123          | 22           |
| CCY_071  | [G/T] | 9        | 0.754  | 0.821  | 0.067 | 34           | 45           | 151          | 13           | 77           | 197          |
| CCY_093  | [G/T] | 10       | 0.413  | 0.478  | 0.065 | 74           | 121          | 34           | 81           | 140          | 68           |
| CCY_202  | [G/A] | 11       | 0.355  | 0.285  | 0.07  | 103          | 106          | 33           | 148          | 123          | 22           |
| CCY_205  | [A/G] | 12       | 0.631  | 0.704  | 0.073 | 33           | 108          | 95           | 21           | 123          | 135          |

[20] 표 1 (계속)

[21]

| 연관: 카이제곱 (df=2) |                  | odds ratio:<br>다중모델 |      |                | HWE 상태      |            | call_rate     |               |
|-----------------|------------------|---------------------|------|----------------|-------------|------------|---------------|---------------|
| Chi_value       | Chi_exact_pValue | 위험<br>대립인           | OR   | CI             | cas_HW      | cas_HW     | cas_call_rate | con_call_rate |
| 5.287           | 3.20E-02         | A1 A                | 2.06 | (1.085, 3.9)   | .174, HWE   | .067, HWE  | 0.99          | 1             |
| 6.041           | 4.88E-02         | A1 A                | 1.37 | (1.055, 1.785) | .569, HWE   | .127, HWE  | 1             | 0.98          |
| 7.299           | 2.60E-02         | A2 C                | 1.41 | (1.085, 1.812) | 1.584, HWE  | 2.407, HWE | 0.92          | 0.99          |
| 6.155           | 4.61E-02         | A1 G                | 1.36 | (1.044, 1.774) | .657, HWE   | .419, HWE  | 0.99          | 0.97          |
| 10.754          | 4.62E-03         | A2 G                | 1.52 | (1.182, 1.961) | .195, HWE   | .029, HWE  | 0.87          | 0.94          |
| 27.984          | 8.38E-07         | A2 T                | 1.49 | (1.156, 1.946) | 44.441, HWD | .075, HWE  | 0.87          | 0.98          |
| 7.34            | 2.55E-02         | A2 G                | 1.43 | (1.103, 1.832) | .945, HWE   | .071, HWE  | 0.9           | 0.98          |
| 14.733          | 6.32E-04         | A2 C                | 1.52 | (1.164, 1.965) | 9.12, HWD   | .185, HWE  | 0.88          | 0.99          |
| 17.789          | 1.37E-04         | A1 G                | 1.49 | (1.102, 2.012) | 56.139, HWD | 2.444, HWE | 0.9           | 0.97          |
| 6.166           | 4.58E-02         | A1 G                | 1.30 | (1.016, 1.666) | 1.747, HWE  | .287, HWE  | 0.89          | 0.98          |
| 6.729           | 3.46E-02         | A2 A                | 1.39 | (1.068, 1.792) | .535, HWE   | .185, HWE  | 0.95          | 0.99          |
| 7.056           | 2.94E-02         | A1 A                | 1.39 | (1.071, 1.805) | .083, HWE   | .863, HWE  | 0.92          | 0.94          |

표 2. 서열번호 1 내지 12의 다형성 서열의 특성

[22]

| ASSAY_ID | rs         | 염색체 | 위치        | 밴드      | 유전자       | 설명                                      | SNP<br>역할                  | 아미노산<br>변화 |
|----------|------------|-----|-----------|---------|-----------|---|----------------------------|------------|
| CCK048   | rs2863383  | 3   | 167396140 | 3q26.1  | 유전자<br>사이 | -                                       | 유전자<br>사이                  | 변화 없음      |
| CCK061   | rs7151139  | 14  | 21933597  | 14q11.2 | C14orf120 | 염색체 14<br>orf 120                       | 인트론                        | 변화 없음      |
| CCK117   | rs724454   | 4   | 91421840  | 4q22.1  | 유전자<br>사이 | -                                       | 유전자<br>사이                  | 변화 없음      |
| CCK162   | rs10142383 | 14  | 21932663  | 14q11.2 | C14orf120 | 염색체 14<br>orf 120                       | 인트론                        | 변화 없음      |
| CCY_041  | rs1402026  | 5   | 114159705 | 5q22.3  | 유전자<br>사이 | -                                       | 유전자<br>사이                  | 변화 없음      |
| CCY_056  | rs1485217  | 3   | 3917830   | 3p26.2  | 유전자<br>사이 | -                                       | 유전자<br>사이                  | 변화 없음      |
| CCY_065  | rs1996489  | 3   | 167399578 | 3q26.1  | 유전자<br>사이 | -                                       | 유전자<br>사이                  | 변화 없음      |
| CCY_067  | rs2236261  | 14  | 21934642  | 14q11.2 | C14orf120 | 염색체 14<br>orf 120                       | coding-synon,<br>reference | 변화 없음      |
| CCY_071  | rs1340655  | 10  | 61325850  | 10q21.2 | ANK3      | ankyrin 3,<br>Ranvier 노드<br>(ankyrin G) | 인트론                        | 변화 없음      |
| CCY_093  | rs1334856  | 13  | 82206386  | 13q31.1 | 유전자<br>사이 | -                                       | 유전자<br>사이                  | 변화 없음      |
| CCY_202  | rs6573195  | 14  | 21934148  | 14q11.2 | C14orf120 | 염색체 14<br>orf 120                       | 인트론                        | 변화 없음      |
| CCY_205  | rs2295706  | 14  | 21935494  | 14q11.2 | C14orf120 | 염색체 14<br>orf 120                       | 인트론                        | 변화 없음      |

[23] 표 1 및 2에 있어서, 각 칼럼이 의미하는 바는 다음과 같다.

[24] · Assay\_ID는 마커의 명칭을 나타내는 것이다.

[25] · SNP는 단일염기 다형성 부위에서 관찰되는 염기를 나타낸 것으로, 여기서 A1과 A2는 시퀀싱 (Sequenom)사의 균질적 MassEXTEND (homogeneous MassEXTEND : 이하 hME라고도 한다) 기법에 의하여 서열분석을 하는 과정에서 질량이 작은 대립인자를 A1 및 질량이 큰 대립인자를 A2로 임의적으로 실험의 편의상 명명한 것이다.

- [26] · 서열번호는 각 SNP 부위를 포함하는 서열의 서열번호를 나타낸 것으로, 각각 101번째 위치에 A1 및 A2 대립인자를 포함하는 서열을 나타낸 것이다.
- [27] · 대립 인자 빈도 (allele frequency) 란에서 cas\_A2, con\_A2 및 Delta는 각각 질병군 (case sample)에서의 A2 대립 인자의 빈도, 정상군 (normal sample)에서의 A2 대립 인자의 빈도 및 상기 cas\_A2와 con\_A2의 차이의 절대값을 나타내는 것이다. 여기서 cas\_A2는 질병군에서  $(A2A2 \text{ 유전자형의 빈도} \times 2 + A1A2 \text{ 유전자형의 빈도}) / (\text{질병군의 샘플 수} \times 2)$ 이고, con\_A2는 정상군에서  $(A2A2 \text{ 유전자형의 빈도} \times 2 + A1A2 \text{ 유전자형의 빈도}) / (\text{정상군의 샘플 수} \times 2)$ 이다.
- [28] · 유전형의 빈도 (genotype frequency)는 각 유전자형의 빈도를 나타내는 것으로, cas\_A1A1, cas\_A1A2 및 cas\_A2A2 및 con\_A1A1, con\_A1A2 및 con\_A2A2는 각각 질병군과 정상군에서 A1A1, A1A2 및 A2A2 유전자형을 갖는 사람의 수를 나타낸다.
- [29] · 카이제곱 (df=2)은 자유도 (degree of freedom)가 2인 카이제곱 값을 나타낸 것으로, Chi-value는 카이제곱 결과 값으로 p-value 계산의 기준이 되는 값이다. chi-exact-p-value는 p-value of Fisher's exact test of chi-square test를 나타내는 것으로, 유전자형 수의 값이 5 보다 작은 값이 포함되는 경우 일반 카이 제곱 검정 결과가 부정확할 수 있으므로, Fisher's exact test를 통해 보다 정확한 통계적 유의성 (p-value)을 검정하는데 사용되는 변수이다. 본 발명에서는  $p\text{-value} \leq 0.05$ 인 경우 질병군과 정상군의 유전자형이 같지 않다 즉, 유의하다고 판단하였다.
- [30] · 위험 대립인자 (risk allele)는 기준 대립인자를 A2로 하여 질병군에서의 A2의 빈도가 정상군에서의 A2의 빈도 보다 큰 경우 (즉,  $\text{cas\_A2} > \text{con\_A2}$ )에는 A2를 위험 대립인자로 하고, 그 반대의 경우에는 A1을 위험 대립인자로 하였다.
- [31] · 오즈 비율 (odds ratio)는 정상군 중에서 위험 대립인자를 가질 확률에 대한 질병군 중에서 위험 대립인자를 가질 확률의 비율을 나타낸다. 본 발명에서는 Mantel-Haenszel odds ratio 방법을 사용하였다. CI는 오즈 비율의 95% 신뢰구간을 나타내는 것으로, (하한 신뢰구간, 상한 신뢰구간)을 나타낸다. 신뢰구간은 1을 포함하는 경우에는 질병과의 연관성을 유의하다고 판단할 수 없다.
- [32] · HWE 상태는 하아디-와인버그 평형 (Hardy-Weinberg Equilibrium)의 상태를 나타내는 것으로, con\_HWE 및 cas\_HWE는 정상군 및 질병군에서 하아디-와인버그 평형 여부를 나타내는 것이다. 카이제곱 (df=1) 검정에서  $\text{chi-value} = 6.63$  ( $p\text{-value} = 0.01$ ,  $\text{df} = 1$ )을 기준으로, 6.63 보다 큰 경우에는 하아디-와인버그 비평형 HWD (Hardy-Weinberg Disequilibrium)으로, 6.63 보다 작은 경우에는 하아디-와인버그 평형 HWE (Hardy-Weinberg Equilibrium)으로 판단하였다.

- [33]       · 콜 비율 (call\_rate)은 원래 실험에 투입한 전체 시료 수에 대한 유전자형이 성공적으로 실험된 시료의 수를 나타내는 것으로, cas\_call\_rate 및 con\_call\_rate는 각각 질병군 및 정상군 실험에 사용된 총 시료 (300 명) 중 유전자형이 성공적으로 분석된 비율을 나타내는 것이다.
- [34]       · rs 번호는 NCBI dbSNP의 확인번호 (identification number)를 의미한다.
- [35]       표 1 및 2는 2005년 2월 1일 NCBI build 119을 기준으로 각 SNP 마커와 관련된 특성을 기재한 것이다.
- [36]       표 1 및 2에 나타낸 바와 같이, 본 발명의 서열번호 1 내지 12의 다형성 마커는 대립인자 출현빈도의 기대치와 관찰치 사이의  $\chi$  제곱 분석 결과, 95% 신뢰도에서 Chi-exact-p-value 값이 0.0000008~0.049의 범위로서 모두 대조군과 유의하게 달랐으며, 오즈 비율 (odds ratio)의 범위가 1.30~2.06으로서 대장암과 연관성이 있다는 것을 알 수 있다.
- [37]       따라서, 본 발명의 폴리뉴클레오티드는 대장암과 관련되는 진단, 유전자 지문 분석 또는 치료에 효과적으로 이용될 수 있다. 구체적으로는 대장암의 진단을 위한 프라이머 또는 프로브로서 사용될 수 있다. 또한, 대장암 치료를 위한 안티센스 DNA 또는 조성물로서 사용될 수 있다.
- [38]       본 발명은 또한, 다형성 부위를 포함한 서열번호 1 내지 12로부터 유래한 10 개 이상의 연속 뉴클레오티드를 포함하는 폴리뉴클레오티드 또는 그 상보체와 혼성화하는 대장암 진단용 대립형질 특이적 폴리뉴클레오티드를 제공한다.
- [39]       상기 대립형질 특이적 (allele-specific) 폴리뉴클레오티드란 각 대립형질에 특이적으로 혼성화하는 것을 의미한다. 즉, 서열번호 1 내지 12의 각 다형성 서열 중의 다형성 부위의 염기를 특이적으로 구별할 수 있도록 혼성화하는 것을 말한다. 여기서, 혼성화란 엄격한 조건, 예를 들면 1M 이하의 염 농도 및 25 °C 이상의 온도하에서 보통 수행된다. 예를 들면, 5xSSPE (750 mM NaCl, 50 mM Na Phosphate, 5 mM EDTA, pH 7.4) 및 25~30°C의 조건이 대립형질 특이적 프로브 혼성화에 적합할 수 있다.
- [40]       본 발명에 있어서 상기 대립형질 특이적 폴리뉴클레오티드는 프라이머일 수 있다. 여기서 "프라이머 (primer)"란 적절한 버퍼 중의 적절한 조건 (예를 들면, 4 개의 다른 뉴클레오시드 트리포스페이트 및 DNA, RNA 폴리머라제 또는 역전사 효소와 같은 중합제) 및 적당한 온도하에서 주형-지시 DNA 합성의 시작점으로서 작용할 수 있는 단일가닥 올리뉴클레오티드를 말한다. 상기 프라이머의 적절한 길이는 사용 목적에 따라 달라질 수 있으나, 통상 15 내지 30 뉴클레오티드이다. 짧은 프라이머 분자는 일반적으로 주형과 안정한 혼성체를 형성하기 위해서는 더 낮은 온도를 필요로 한다. 프라이머 서열은 주형과 완전하게 상보적일 필요는 없으나, 주형과 혼성화할 정도로 충분히

상보적이어야 한다. 상기 프라이머는 그 3'말단이 서열번호 1 내지 12의 다형성 부위 (염기 n)와 정렬하는 것이 바람직하다. 상기 프라이머는 다형성 부위를 포함하는 표적 DNA에 혼성화하고, 상기 프라이머가 완전한 상동성을 보이는 대립형질 형태의 증폭을 개시한다. 이 프라이머는 반대편에 혼성화하는 제2 프라이머와 쌍을 이루어 사용된다. 증폭에 의하여 두개의 프라이머로부터 산물이 증폭되고, 이는 특정 대립형질 형태가 존재한다는 것을 의미한다. 본 발명의 프라이머에는 리가제 연쇄 반응 (ligase chain reaction : LCR)에서는 사용되는 폴리뉴클레오티드 단편을 포함한다.

- [41] 본 발명에 있어서, 상기 대립형질 특이적 폴리뉴클레오티드는 프로브일 수 있다. 본 발명에서 "프로브 (probe)"란 혼성화 프로브를 의미하는 것으로, 핵산의 상보성 가닥에 서열 특이적으로 결합할 수 있는 올리고뉴클레오티드를 의미한다.
- [42] 이러한 프로브에는 Nielsen 등, Science 254, 1497-1500 (1991)에 기재된 펩티드 핵산을 포함한다. 본 발명의 프로브는 대립형질 특이적 프로브로서, 같은 종의 두 구성원으로부터 유래한 핵산 단편 중에 다형성 부위가 존재하여, 한 구성원으로부터 유래한 DNA 단편에는 혼성화하나, 다른 구성원으로부터 유래한 단편에는 혼성화 하지 않는다. 이 경우 혼성화 조건은 대립형질간의 혼성화 강도에 있어서 유의한 차이를 보여, 대립형질 중 하나에만 혼성화하도록 충분히 엄격해야 한다. 이러한 본 발명의 프로브는 중앙부위 (즉, 15 개의 뉴클레오티드로 된 프로브이면 7번 위치가, 16 개의 뉴클레오티드로 된 프로브이면 8 또는 9번 위치)가 상기 서열의 다형성 부위와 정렬하는 것이 바람직하다. 이렇게 함으로써 다른 대립형질성 형태 간에 좋은 혼성화 차이를 유발할 수 있다. 본 발명의 상기 프로브는 대립형질을 검출 하기 위한 진단 방법 등에 사용될 수 있다. 상기 진단 방법에는 서던 블롯팅 등과 같은 핵산의 혼성화에 근거한 검출방법들이 포함되며, DNA 칩을 이용한 방법에서 DNA 칩의 기판에 미리 결합된 형태로 제공될 수도 있다.
- [43] 본 발명은 또한, 본 발명의 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드를 포함하는 대장암 진단용 마이크로어레이를 포함한다. 상기 마이크로어레이는 DNA 또는 RNA 폴리뉴클레오티드를 포함하는 것일 수 있다. 상기 마이크로어레이는 본 발명의 폴리뉴클레오티드를 포함하는 것을 제외하고는 통상적인 마이크로 어레이로 이루어진다.
- [44] 본 발명은 또한, 본 발명의 폴리뉴클레오티드를 포함하는 대장암 진단용 키트를 제공한다. 상기 키트에는 본 발명의 폴리뉴클레오티드 뿐만 아니라, 중합 반응에 필요한 시약, 예를 들면 dNTP, 각 종의 중합효소 및 발색제 등을 포함할 수 있다.



- [45] 또한, 본 발명은 피검체로부터 핵산 시료를 얻는 단계; 및
- [46] 서열번호 1 내지 12의 폴리뉴클레오티드 또는 그 상보적 폴리뉴클레오티드 중의 하나 이상의 101번 위치의 다형성 부위 (n)의 뉴클레오티드 서열을 결정하는 단계를 포함하는 대장암 진단 방법을 제공한다. 여기서, 상기 다형성 부위의 뉴클레오티드 서열이 상기 표 1 및 2에 나타낸 바와 같이, 위험대립인자인 경우 피검체는 대장암 환자이거나 대장암에 걸릴 확률이 높은 것으로 판단하는 단계를 포함할 수 있다.
- [47] 상기 방법 중 먼저 피검체로부터 핵산을 얻는 단계는 통상의 DNA 분리방법에 의하여 수행될 수 있다. 예를 들면, 표적 핵산을 PCR을 통하여 증폭하고 이를 정제하여 얻을 수 있다. 그 외 리가제 연쇄 반응 (LCR) (Wu 및 Wallace, Genomics 4, 560 (1989), Landegren 등, Science 241, 1077 (1988)), 전사증폭 (transcription amplification) (Kwoh 등, Proc. Natl. Acad. Sci. USA 86, 1173 (1989)) 및 자가 유지 서열 복제 (Guatelli 등, Proc. Natl. Acad. Sci. USA 87, 1874 (1990)) 및 핵산에 근거한 서열 증폭 (NASBA)이 사용될 수 있다. 마지막 두 가지의 방법은 등온전사에 근거한 등온 반응과 관련되는 것으로서, 증폭산물로서 30 또는 100배의 단일가닥 RNA 및 이중 가닥 DNA를 생산한다.
- [48] 본 발명의 방법의 일 구체예는, 상기 다형성 부위의 뉴클레오티드 서열을 결정하는 단계는 본 발명의 서열번호 1 내지 12로 구성된 군으로부터 유래한 10 개 이상의 연속 뉴클레오티드를 포함하는 폴리뉴클레오티드로서, 101 번째 염기를 포함하는 대장암 진단 또는 치료용 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드가 고정화되어 있는 마이크로어레이에 상기 핵산 시료를 혼성화시키는 단계; 및 상기 혼성화 결과를 검출하는 단계를 포함하는 것일 수 있다.
- [49] 마이크로어레이 및 프로브 폴리뉴클레오티드를 기판 상에 고정화하여 마이크로어레이를 제조하는 방법은 당업계에 잘 알려져 있다. 본 발명의 대장암과 연관된 프로브 폴리뉴클레오티드의 기판 상에 고정화하는 과정도 또한 이러한 종래 기술을 사용하여 용이하게 제조할 수 있다. 또한, 마이크로어레이 상에서의 핵산의 혼성화 및 혼성화 결과의 검출은 당업계에 잘 알려져 있다. 상기 검출은 예를 들면, 핵산 시료를 형광 물질 예를 들면 Cy3 및 Cy5와 같은 물질을 포함하는 검출가능한 신호를 발생시킬 수 있는 표지 물질로 표지한 다음, 마이크로어레이 상에 혼성화하고 상기 표지 물질로부터 발생하는 신호를 검출함으로써 혼성화 결과를 검출할 수 있다.

#### 유리한 효과

- [50] 본 발명의 폴리뉴클레오티드에 의하면, 대장암의 진단, 치료 및 유전자 지문 분석과 같은 대장암과 관련된 용도로 사용될 수 있다.

- [51] 또한, 본 발명의 폴리뉴클레오티드를 포함하는 마이크로어레이 및 키트에 의하면, 대장암을 효과적으로 진단할 수 있다.
- [52] 본 발명의 대장암과 관련되는 폴리뉴클레오티드를 분석하는 방법에 의하면, 대장암의 존재 또는 위험의 유무를 효과적으로 진단할 수 있다.
- 발명의 실시를 위한 최선의 형태**
- [53] 이하 본 발명을 실시예를 통하여 보다 상세하게 설명한다. 그러나, 이들 실시예는 본 발명을 예시적으로 설명하기 위한 것으로 본 발명의 범위가 이들 실시예에 한정되는 것은 아니다.
- [54] 실시예
- [55] 실시예1
- [56] 본 실시예에서는 한국인 중 대장암 환자로 판명되어 치료 중인 환자군 (300 명)과 환자군과 동일 연령 대에 해당하면서 아직 대장암의 증상이 없는 정상인 (300 명)의 혈액으로부터 DNA를 분리하고, 특정한 단일염기 다형의 출현 빈도를 분석하였다. 본 실시예에 선택된 단일염기 다형은 공개된 데이터베이스 (NCBI dbSNP: <http://www.ncbi.nlm.nih.gov/SNP/>) 또는 Sequenom사의 [realsnp.com](http://www.realsnp.com/) (<http://www.realsnp.com/>)로부터 선택하였으며, 선택된 단일염기 다형 주변의 서열을 이용한 프라이머를 이용하여 시료 중의 단일염기 서열을 분석하였다.
- [57] 1. DNA 시료의 준비
- [58] 대장암 환자 및 정상인으로부터 수집한 혈액으로부터 DNA를 추출하였다. DNA 추출은 공지의 추출 방법 (Molecular cloning : A Laboratory Manual, p 392, Sambrook, Fritsch and Maniatis, 2nd edition, Cold Spring Harbor Press, 1989)과 상업용 키트 (Gentra system)의 설명서에 따라 이루어졌다. 추출된 DNA 중 순도가 UV 측정치 (260/280nm)가 최소 1.6 이상되는 것만을 선별하여 사용하였다.
- [59] 2. 표적 DNA의 증폭
- [60] 분석하고자 하는 단일염기 다형이 포함된 일정한 DNA 영역인 표적 DNA를 PCR을 이용하여 증폭하였다. PCR 방법은 통상적인 방법으로 진행하였으며, 그 조건은 다음과 같았다. 먼저, 표적 게놈 DNA를 2.5 ng/ml로 준비하였다. 다음으로 다음의 PCR 반응액을 준비하였다.
- [61] 물(HPLC 급) 2.24 $\mu$ l
- [62] 10x 버퍼 (15 mM MgCl<sub>2</sub>, 25 mM MgCl<sub>2</sub> 함유) 0.5 $\mu$ l
- [63] dNTP 믹스(GIBCO)(25 mM/각) 0.04 $\mu$ l
- [64] Taq pol(HotStar)(5U/ $\mu$ l) 0.02 $\mu$ l
- [65] 전위/후위 프라이머 믹스 (1 $\mu$ M/각) 0.02 $\mu$ l
- [66] DNA 1.00 $\mu$ l

- [67] 총 반응 부피  $5.00\mu\text{l}$
- [68] 여기서, 상기 전위(forward) 및 후위(reverse) 프라이머는 공지의 데이데이터 베이스의 단일염기 다형의 상류와 하류로부터, 적당한 위치에서 선택하였다. 상기 프라이머를 표 3에 정리하였다.
- [69] 열 순환 반응은,  $95^{\circ}\text{C}$ 에서 15분 동안 유지하고,  $95^{\circ}\text{C}$ 에서 30초,  $56^{\circ}\text{C}$ 에서 30초,  $72^{\circ}\text{C}$ 에서 1분을 45회 반복하고,  $72^{\circ}\text{C}$ 에서 3분 동안 유지한 후,  $4^{\circ}\text{C}$ 에 보관하였다. 그 결과, 200 개 뉴클레오티드 이하의 길이를 가진 표적 DNA 단편을 얻었다.
- [70] 3. 증폭된 표적 DNA 중의 단일염기 다형의 분석
- [71] 표적 DNA 단편 중의 단일염기 다형의 분석은 시퀀넘 (Sequenom)사의 균질적 MassEXTEND (homogeneous MassEXTEND : 이하 hME라고도 한다) 기법을 이용하였다.
- [72] 상기 MassEXTEND 기법의 원리는 다음과 같다. 먼저, 표적 DNA 단편 중의 단일염기 다형 바로 전까지의 염기에 상보적인 프라이머 (연장용 프라이머 (extension primer)라고도 한다.)를 제작한다. 다음으로, 상기 프라이머를 표적 DNA 단편에 혼성화시키고, DNA 중합 반응을 일으킨다. 이 때, 반응액 중에 대상 단일염기 다형 대립인자 중 제1 대립인자 염기 (예를 들면, A 대립인자)에 상보적인 염기가 첨가된 후 반응이 멈추도록 하는 시약 (예를 들면, ddTTP)을 포함시킨다. 그 결과, 표적 DNA 단편에 제1 대립인자 (예를 들면, A 대립인자)가 존재하는 경우에는, 상기 제1 대립인자에 상보적인 염기 (예를 들면, T) 하나만이 첨가된 산물이 얻어진다. 반면, 표적 DNA 단편에 제2 대립인자 (예를 들면, G 대립인자)가 존재하는 경우에는, 상기 제2 대립인자에 상보적인 염기 (예를 들면, C)가 첨가된 가장 근접한 제1 대립인자 염기 (예를 들면, A)까지 연장된 산물을 얻게 된다. 이렇게 얻어진 상기 프라이머로부터 연장된 산물의 길이를 질량 분석을 통하여 결정함으로써, 표적 DNA에 존재하는 대립인자의 종류를 결정할 수 있다. 구체적인 실험조건은 다음과 같았다.
- [73] 먼저, 상기 PCR 산물로부터 결합되지 않는 dNTP를 제거하였다. 이를 위하여 순수  $1.53\mu\text{l}$ , HME 버퍼  $0.17\mu\text{l}$ , SAP (shrimp alkaline phosphatase)  $0.30\mu\text{l}$ 를 1.5 ml 튜브에 넣고 혼합하여 SAP 효소 용액을 준비하였다. 상기 튜브를 5,000 rpm에서 10 초 동안 원심분리하였다. 그 후, PCR 산물을 상기 SAP 용액 튜브에 넣고, 밀봉한 다음  $37^{\circ}\text{C}$ 에서 20분,  $85^{\circ}\text{C}$ 에서 5분 동안 유지한 후,  $4^{\circ}\text{C}$ 에 보관하였다.
- [74] 다음으로, 상기 표적 DNA 산물을 주형으로 하여 균질적 연장 반응을 실시하였다. 반응액은 다음과 같았다.
- [75] 물(나노급 순수)  $1.728\mu\text{l}$
- [76] hME 연장 믹스 (2.25 mM d/ddNTPs를 포함하는 10x버퍼)  $0.200\mu\text{l}$

[77] 연장 프라이머 (각 100  $\mu\text{M}$ ) 0.054  $\mu\text{l}$

[78] Thermosequenase(32U/ $\mu\text{l}$ ) 0.018  $\mu\text{l}$

[79] 총부피 2.00  $\mu\text{l}$

[80] 상기 반응액을 잘 혼합한 후, 스피다운 원심분리하였다. 상기 반응액이 든 튜브 또는 플레이트를 잘 밀봉한 다음, 94  $^{\circ}\text{C}$ 에서 2분 동안 유지한 다음, 94  $^{\circ}\text{C}$ 에서 5초, 52  $^{\circ}\text{C}$ 에서 5초, 72  $^{\circ}\text{C}$ 에서 5초를 40회 반복 한 다음, 4  $^{\circ}\text{C}$ 에 보관하였다. 이렇게 얻어진 균질적 연장 반응 산물을 레진 (SpectroCLEAN)을 사용하여 세척하였다. 연장 반응에 사용된 연장 프라이머는 표 3에 나타내었다.

[81] 표 3. 표적 DNA 증폭에 사용된 프라이머 및 균질 연장 반응에 사용된 연장 프라이머

[82]

| 마커<br>명칭 | 표적 DNA 증폭용 프라이머<br>(서열번호) |         | 연장 프라이머<br>(서열번호) |
|----------|---------------------------|---------|-------------------|
|          | 전위 프라이머                   | 후위 프라이머 |                   |
| CCK048   | 13                        | 14      | 15                |
| CCK061   | 16                        | 17      | 18                |
| CCK117   | 19                        | 20      | 21                |
| CCK162   | 22                        | 23      | 24                |
| CCY_041  | 25                        | 26      | 27                |
| CCY_056  | 28                        | 29      | 30                |
| CCY_065  | 31                        | 32      | 33                |
| CCY_067  | 34                        | 35      | 36                |
| CCY_071  | 37                        | 38      | 39                |
| CCY_093  | 40                        | 41      | 42                |
| CCY_202  | 43                        | 44      | 45                |
| CCY_205  | 46                        | 47      | 48                |

[83] 얻어진 연장 반응 산물을 질량 분석 방법 중 MALDI-TOF (Matrix Assisted Laser Desorption and Ionization-Time of Flight)를 이용하여 다형성 부위의 서열을 분석하였다. 상기 MALDI-TOF는 분석하고자 하는 물질이 레이저 빔을 받으면, 이온화된 매트릭스 (3-Hydroxypicolinic acid)와 함께 비행하여 진공상태에서 반대편에 있는 검출기까지 날아가는 데 걸린 시간을 계산하여 질량을 분석해내는 원리에 의하여 작동한다. 질량이 작은 물질은 검출기에 빨리 도달하게 되는데, 이렇게 얻어지는 질량의 차이와 이미 알고 있는 단일염기 다형의 서열을 근거로 하여 표적 DNA의 단일염기 다형의 서열을 결정할 수 있는 것이다.

[84] 상기과 같은 MALDI-TOF를 사용한 표적 DNA의 다형성 부위의 서열을 결정한 결과는 표 1 및 2에 나타내었다. 이때, 각 대립인자는 각 개체에 있어서 동형접합자 (homozygote) 또는 이형접합자 (heterozygote)의 형태로 존재할 수 있다. 멘델의 유전법칙과 하디-와인버그 (Hardy-Weinberg) 법칙에 의하면,

집단을 구성하는 대립인자의 조성은 일정한 빈도로 유지되며, 통계적으로 유의한 경우 생물학적 기능상의 의미를 부여할 수 있다. 본 발명의 단일염기 다형은 표 1 및 2에 나타낸 바와 같이, 대장암 환자에게서 통계적으로 유의한 수준으로 출현되어, 대장암의 진단 등에 사용될 수 있음을 알 수 있었다.

## 청구의 범위

- [1] 서열번호 1 내지 12로 구성된 군으로부터 유래한 10 개 이상의 연속 뉴클레오티드를 포함하는 폴리뉴클레오티드로서, 101 번째 염기를 포함하는 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드.
- [2] 제1항의 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드와 혼성화하는 폴리뉴클레오티드.
- [3] 제1항 또는 제2항에 있어서, 길이가 10 내지 100 뉴클레오티드인 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드.
- [4] 제1항에 있어서, 프라이머 또는 프로브인 폴리뉴클레오티드.
- [5] 제1항의 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드를 포함하는 대장암 진단용 마이크로어레이.
- [6] 제1항의 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드를 포함하는 대장암 진단용 키트.
- [7] 피검체로부터 핵산 시료를 얻는 단계; 및  
서열번호 1 내지 12의 폴리뉴클레오티드 또는 그 상보적 폴리뉴클레오티드 중의 하나 이상의 다형성 부위 (101 번째)의 뉴클레오티드 서열을 결정하는 단계를 포함하는 대장암 진단 방법.
- [8] 제7항에 있어서, 상기 다형성 부위의 뉴클레오티드 서열을 결정하는 단계는 제1항에 따른 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드가 고정화되어 있는 마이크로어레이에 상기 핵산 시료를 혼성화시키는 단계; 및  
상기 혼성화 결과를 검출하는 단계를 포함하는 것인 방법.
- [9] 제7항에 있어서, 상기 다형성 부위의 뉴클레오티드 서열을 결정한 결과 서열 번호 1 내지 12의 다형성 부위의 뉴클레오티드가 각각의 위험 대립인자인 A,A,C,G,G,T,G,C,G,G,A 및 A로 이루어지는 군으로부터 선택되는 하나 이상의 뉴클레오티드가 존재하는 경우 상기 피검체를 대장암에 걸릴 확률이 높은 위험군에 속하는 것으로 판정하는 단계를 포함하는 방법.

## 요약서

본 발명은 서열번호 1 내지 12로 구성된 군으로부터 유래한 10 개 이상의 연속 뉴클레오티드를 포함하는 폴리뉴클레오티드로서, 101 번째 염기를 포함하는 폴리뉴클레오티드 또는 그의 상보적 폴리뉴클레오티드를 제공한다.